Cloudera

CDP-ADMIN-PVC Exam

Cloudera Administrator on premises Certification Exam

Exam Latest Version: 6.0

DEMO Version

Full Version Features:

- 90 Days Free Updates

- 30 Days Money Back Guarantee

- Instant Download Once Purchased

- 24 Hours Live Chat Support

Full version is available at link below with affordable price.

https://www.directcertify.com/cloudera/cdp-admin-pvc

Which of the following components are essential for the functioning of the Cloudera Data Platform (CDP) on-premises Data Lake, focusing on data storage and metadata management?

A: HDFS (Hadoop Distributed File System) for data storage.
B: Hive Metastore for metadata management.
C: YARN (Yet Another Resource Negotiator) for resource management.
D: ZooKeeper for coordinating distributed processes.
E: Kafka for real-time data streaming.

Correct Answer: A, B, C, D

Explanation:

HDFS provides distributed storage, Hive Metastore manages metadata, YARN manages resources and ZooKeeper coordinates distributed processes are all essential for a functioning Data Lake. Kafka is more related to streaming ingest rather than core Data Lake functionality, though it may be part of the overall architecture. The question specifically asks about Data Lake functioning.

Question 2. (Single Select)

You are configuring resource pools in YARN for different departments in your organization. The marketing department requires a guaranteed minimum of 40% of the cluster's resources, while the engineering department requires at least 50%. What is the most appropriate configuration for the 'yarn.scheduler.capacity.maximum-am-resource-percent' property in 'yarn-site.xmP , considering efficient resource utilization and potentially variable workloads?

A: Set 'yarn.scheduler.capacity.maximum-am-resource-percent' to 0.1 (10%).
B: Set 'yarn.scheduler.capacity.maximum-am-resource-percent' to 0.4 (40%).
C: Set 'yarn.scheduler.capacity.maximum-am-resource-percent' to 0.5 (50%).
D: Set 'yarn.scheduler.capacity.maximum-am-resource-percent' to 0.9 (90%).
E: Set 'yarn.scheduler.capacity.maximum-am-resource-percent' to 1.0 (100%).

Correct Answer: D

Explanation:

'yarn.scheduler.capacity.maximum-am-resource-percent' defines the maximum percentage of resources in the cluster that can be used for ApplicationMasters (AMs). While marketing and engineering require 40% and 50% respectively of overall cluster resources, the AMs need resources too. The maximum usage percentage should allow enough space for the Application Master processes to start for a resource intensive job. Setting this to a lower value might lead to resource starvation for the application masters.

Question 3. (Multi Select)

Your Cloudera Data Platform (CDP) on-premises cluster is experiencing frequent HDFS NameNode failures, leading to data unavailability. You want to implement High Availability (HA) for the NameNode. What steps are essential to configure NameNode HA using Quorum Journal Manager (QJM)?

A: Install and configure a single ZooKeeper server.
B: Configure two NameNode hosts and share a single edit log directory between them.
C: Install and configure a quorum of JournalNodes (typically 3 or more) to store the edit log.
D: Configure the 'dfs.nameservices' and 'dfs.ha.namenodes.[nameserviceId]' properties in 'hdfs-site.xmr.
E: Format the NameNode using the '-format' option on both NameNode hosts.

Correct Answer: C, D

Explanation:

QJM requires a quorum of JournalNodes to store the edit log redundantly. You also need to configure 'dfs.nameserviceS and dfs.ha.namenodes.[nameserviceId]' to enable HA. A single ZooKeeper server is not enough for HA. NameNodes should not share a single edit log directory. Formatting with -format should be done only on the initial NameNode setup and not on both after configuring HA.

Question 4. (Multi Select)

A user reports that they are unable to access a specific directory in HDFS, even though they believe they have the correct permissions. You need to troubleshoot the access issue. Which of the following actions would be most helpful in diagnosing the problem?

A: Check the HDFS audit logs to see if the user's access attempts are being logged and if any errors are reported.
B: Use the 'hdfs dfs -Is -R command to recursively list the directory and its contents, paying attention to the permissions of each file and directory.
C: Verify the user's group membership and check if any ACLs (Access Control Lists) are configured on the directory or its parent directories.
D: Restart the NameNode to refresh the permissions cache.
E: Increase the HDFS replication factor to improve data availability.

Correct Answer: A, B, C

Explanation:

Checking audit logs provides insights into access attempts and errors. Listing the directory recursively shows the permissions of each file and directory. Verifying group membership and ACLs confirms if the user has the necessary permissions. Restarting the NameNode is unlikely to resolve a permission issue, and increasing the replication factor does not affect access control.

Question 5. (Single Select)

You are planning a CDP on-premises deployment and need to choose the appropriate hardware for your HDFS DataNodes. Your workload consists of large sequential reads and writes. Which storage configuration would provide the best performance and cost- effectiveness?

A: Solid-state drives (SSDs) in a RAID 0 configuration.
B: High-capacity SATA hard disk drives (HDDs) in a RAID 10 configuration.
C: Small-capacity SAS hard disk drives (HDDs) in a RAID 5 configuration.

D: A mix of SSDs for frequently accessed data and HDDs for less frequently accessed data, leveraging HDFS tiering.

E: NVMe drives in JBOD (Just a Bunch of Disks) configuration.

Correct Answer: B

Explanation:

For large sequential reads and writes, high-capacity HDDs in RAID 10 offer a good balance of performance and cost-effectiveness. SSDs are faster but more expensive and may not be necessary for sequential workloads. RAID 5 has write performance limitations. HDFS tiering is a good option for mixed workloads, but the question specifies a sequential workload. NVMe drives in JBOD are very fast, but lack redundancy unless HDFS replication covers it. So RAIDIO configuration of HDDs is the most appropriate answer here.

Full version is available at link below with affordable price.

https://www.directcertify.com/cloudera/cdp-admin-pvc